Cold streams in early massive hot haloes as the main mode of galaxy formation

Dekel, A ; Birnboim, Y ; Engel, G ; Freundlich, J ; Goerdt, T ; Mumcuoglu, M ; Neistein, E ; Pichon, C ; Teyssier, R ; Zinger, E

Abstract: Massive galaxies in the young Universe, ten billion years ago, formed stars at surprising intensities1, 2. Although this is commonly attributed to violent mergers, the properties of many of these galaxies are incompatible with such events, showing gas-rich, clumpy, extended rotating disks not dominated by spheroids1, 2, 3, 4, 5. Cosmological simulations6 and clustering theory6, 7 are used to explore how these galaxies acquired their gas. Here we report that they are 'stream-fed galaxies', formed from steady, narrow, cold gas streams that penetrate the shock-heated media of massive dark matter haloes8, 9. A comparison with the observed abundance of star-forming galaxies implies that most of the input gas must rapidly convert to stars. One-third of the stream mass is in gas clumps leading to mergers of mass ratio greater than 1:10, and the rest is in smoother flows. With a merger duty cycle of 0.1, three-quarters of the galaxies forming stars at a given rate are fed by smooth streams. The rarer, submillimetre galaxies that form stars even more intensely2, 12, 13 are largely merger-induced starbursts. Unlike destructive mergers, the streams are likely to keep the rotating disk configuration intact, although turbulent and broken into giant star-forming clumps that merge into a central spheroid4, 10, 11. This stream-driven scenario for the formation of discs and spheroids is an alternative to the merger picture.

DOI: https://doi.org/10.1038/nature07648

Early Massive Galaxy Formation by Cold Streams Through Hot Haloes

A. Dekel[1], Y. Birnboim[1], G. Engel[1], J. Freundlich[1,2], T. Goerdt[1], M. Mumcuoglu[1], E. Neistein[1], C. Pichon[3], R. Teyssier[4], & E. Zinger[1]

[1]Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

[2]Department de Physique, Ecole Normale Superieure, 24 rue Lhomond, 75231 Paris cedex 05, France

[3]Institut d'Astrophysique de Paris and UPMC, 98bis Boulevard Arago, Paris 75014, France

[4]Institut de Recherches sur les lois Fondamentales de l'Univers, DSM, l'Orme des Merisiers, 91198 Gif-sur-Yvette, France

---

The massive galaxies in the young universe, ten billion years ago, formed stars at surprising intensities.[1,2] While this was commonly attributed to violent mergers, many of these galaxies seem to be extended rotating discs incompatible with mergers.[3,2,4] In order to uncover the origin of this phenomenon, we use a state-of-the-art cosmological simulation[5] and clustering theory[6,7] to explore how these galaxies acquired their gas. We find that these are "Stream-Fed Galaxies", growing via steady, narrow, cold gas streams, which penetrate effectively through the shock-heated media of dark-matter haloes as massive as the Milky Way's. This confirms an earlier conjecture.[8] Half the stream mass is in clumps leading to mergers of mass ratio 1:10 or higher, and half is in smoother flows. Since the merger duty cycle is 0.1, three-quarters of the galaxies forming stars at a given rate are fed by smooth streams. Unlike destructive major mergers, the smoother flows can keep the discs intact, though thick and perturbed. The observed abundance of star-forming galaxies implies that the inflowing gas turns into stars at maximum efficiency. In contrast, the sub-millimeter galaxies that form stars even more intensely[1,9,10] are largely compact merger-induced starbursts in haloes twice as massive.

Star-Formation Rate versus Halo Growth Rate

It appears that the most effective star formers in the universe were galaxies of stellar and gas mass on the order of $10^{11} M_\odot$ at redshifts $z = 2$–3,[11,2] when the universe was about 3 Gyr old. The common cases[3,2] show star-formation rates (SFR) of $100 - 200\, M_\odot\, \mathrm{yr}^{-1}$. These include UV-selected galaxies termed BX/BM, and rest-frame optically selected galaxies termed sBzK, which we jointly refer to as "Star-Forming Galaxies" (SFG). Their SFR is much higher than the $4\, M_\odot\, \mathrm{yr}^{-1}$ in the Milky Way today, while their characteristic

dynamical time is only about 5 times shorter. The comoving space density of SFGs is $n \simeq 2\times10^{-4}\,\mathrm{Mpc}^{-3}$, implying within the standard ΛCDM cosmology that they reside in dark-matter haloes of masses $\lesssim 3.5\times10^{12}M_\odot$. In parallel, the most extreme star formers are observed as dusty objects, termed Sub-Millimeter Galaxies (SMG),[10,9] with SFRs up to $\sim 1,000\,M_\odot\,\mathrm{yr}^{-1}$ and $n \simeq 2\times10^{-5}\,\mathrm{Mpc}^{-3}$. While the SMGs could largely be starbursts induced by major mergers, the morphology and kinematics of the SFGs indicate extended, thick rotating discs that are incompatible with the expected compact or highly perturbed appearance of ongoing major mergers.[3,2,12,4] The big puzzle is how massive galaxies form most of their stars so efficiently at early times and not through major mergers, given that there is no equivalent phenomenon in today's universe and its origin is not obvious theoretically. A necessary condition is clearly a steady, rapid gas supply into appropriately massive discs at early epochs.

One should first verify that the required rate of gas supply is compatible with the predicted growth rate of the corresponding dark-matter haloes in the standard cosmological model. The average growth rate of halo virial mass M_v, via mergers and smooth accretion, is derived[6] based on the EPS theory of gravitational clustering[13] (Methods). It resembles the assembly rate in cosmological N-body simulations.[14,15] For the ΛCDM cosmology (a flat universe with 72% dark energy, mass dominated by cold dark matter, and fluctuation normalization parameter $\sigma_8 = 0.8$),[16] the corresponding growth rate of the baryonic component is well fitted by the practical formula[6]

$$\dot{M} \simeq 6.6\,M_{12}^{1.15}\,(1+z)^{2.25}\,f_{.165}\;M_\odot\,\mathrm{yr}^{-1}\;, \qquad (1)$$

where $M_{12} \equiv M_\mathrm{v}/10^{12}M_\odot$, and $f_{.165}$ is the baryonic fraction in the matter assembled into haloes in units of the cosmological value $f_\mathrm{b}=0.165$. Thus, at $z=2.2$, the average baryonic growth rate of haloes of $2\times10^{12}M_\odot$ is predicted to be $\dot{M}\simeq 200\,M_\odot\,\mathrm{yr}^{-1}$, fairly sufficient for feeding the SFR observed in SFGs. However, this is not by a large margin, implying that (a) the incoming material must be mostly gaseous, (b) the cold gas must very efficiently penetrate deep into the inner halo, and (c) the SFR must closely follow the gas supply rate.

Penetrating Cold Narrow Streams

The required efficient penetration into the inner halo is not a trivial matter, given that the indicated halo masses of $M_\mathrm{v} > 10^{12}M_\odot$ are above the threshold for virial shock heating,[17–19,8] $M_\mathrm{shock} \lesssim 10^{12}M_\odot$. Such haloes are encompassed by a stable shock near their outer radius R_v, inside which gravity and thermal energy are in virial equilibrium. Infalling gas through the virial shock is expected to heat up to the virial temperature and stall in quasi-static equilibrium before it can cool and gradually rain into the inner galaxy.[20] However, Dekel & Birnboim[8] have conjectured that at $z \geq 2$, these hot massive haloes are penetrated by narrow cold streams. The reason is that at early times, the haloes of $M_\mathrm{v} > M_\mathrm{shock}$ populate the extreme massive tail of the halo mass distribution. As such, they are fed by dark-matter filaments from the cosmic web that are narrow compared to R_v and denser than the mean density within the halo. The enhanced density of the gas that streams along these filaments allows it to cool more rapidly than the dynamical compression rate behind

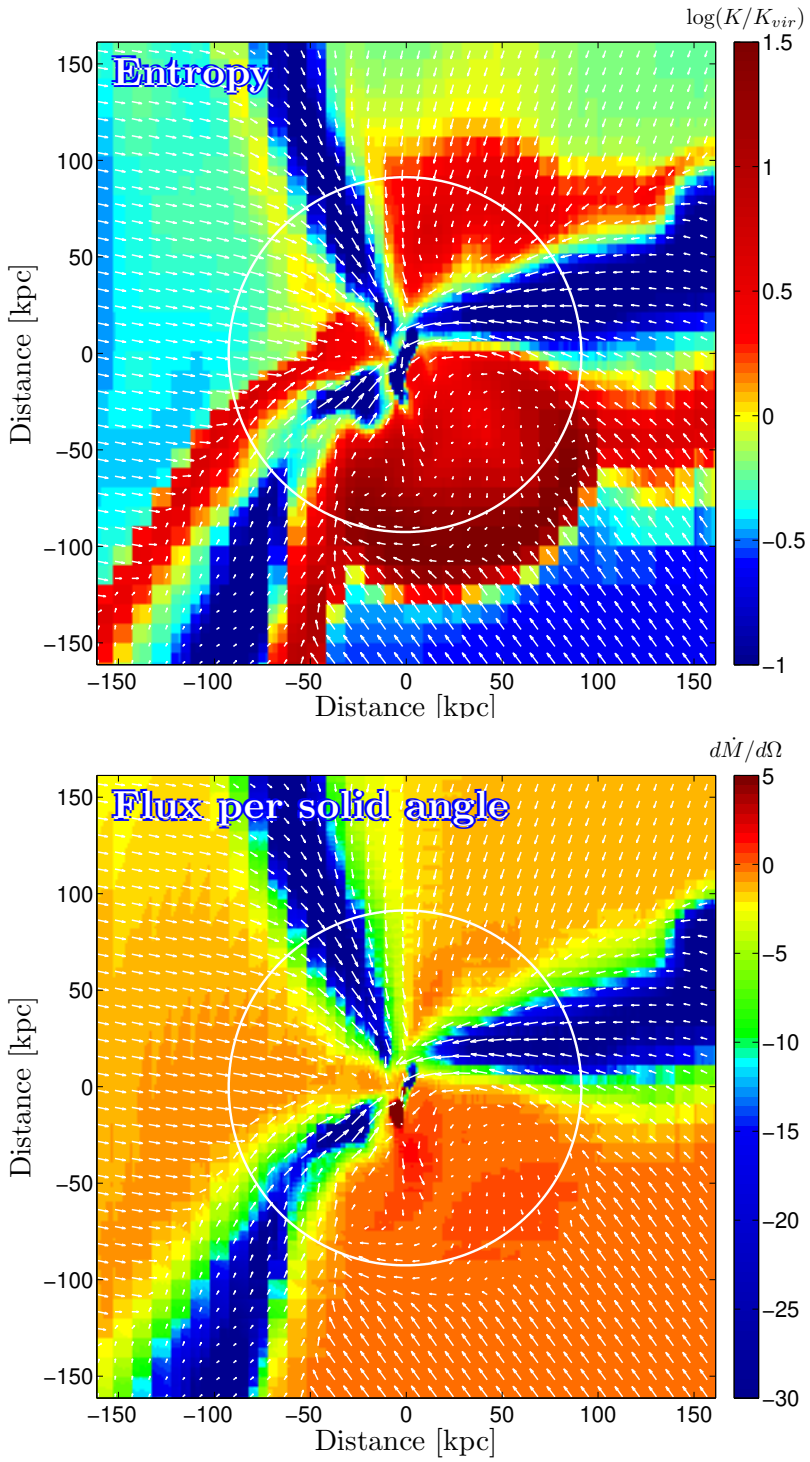

Figure 1. Gas maps in a thin slice through a galaxy of $M_{\rm v}=10^{12}M_\odot$ at $z=2.5$. Arrows describe the velocity field and the circle marks the halo virial radius $R_{\rm v}$. The **entropy**, $\log K=\log(T/\rho^{2/3})$ (in units of the virial quantities), highlights (in red) the high-entropy medium filling the halo out to the virial shock near $R_{\rm v}$. It exhibits three, radial, low-entropy streams (blue) penetrating into the inner disc seen edge-on. The radial **flux** per solid angle, $\dot{m}=r^2\rho\,v_r$ (in $M_\odot\,{\rm yr}^{-1}{\rm rad}^{-2}$), demonstrates that almost all the inward flux is channeled through the streams (blue), at a rate that remains roughly the same at all radii. This rate is several times higher than the spherical average outside the virial sphere, $\dot{m}_{\rm vir}\simeq 8\,M_\odot\,{\rm yr}^{-1}{\rm rad}^{-2}$.

a shock and thus avoid the shock heating that occurs elsewhere in the halo (Supplementary Information).

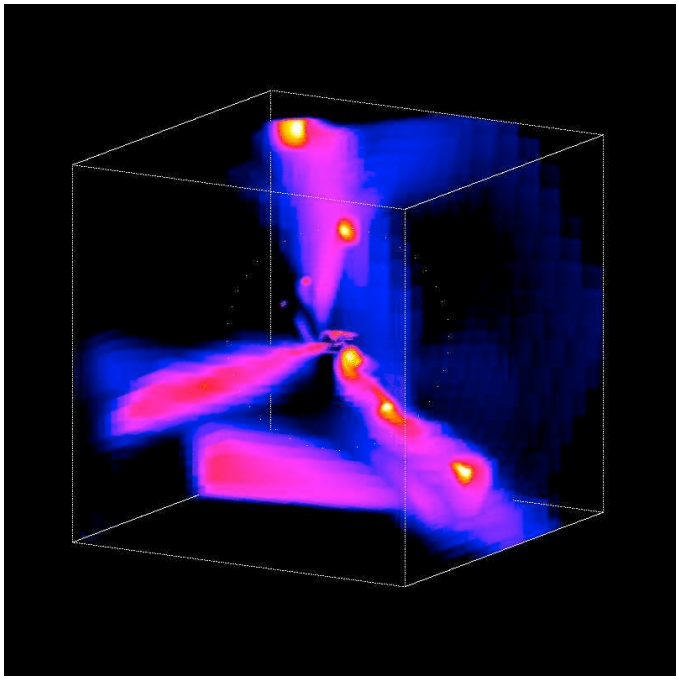

Figure 2. Radial flux map of the galaxy shown in Fig. 1, highlighting the three-dimensional structure of the streams in a box of side 320 kpc. The colors refer to inflow rate per solid angle of point-like tracers at the centers of the cubic grid cells. The dotted circle marks the halo virial radius. Two of the three radial streams show gas clumps of mass on the order of one tenth of the central galaxy, but most of the streams are smoother.

To test this conjecture, we study the way gas feeds massive high-z galaxies in the MareNostrum simulation — a hydrodynamical simulation in a comoving cosmological box of $50\,h^{-1}$Mpc and a resolution of $\sim 1\,h^{-1}$kpc at the galaxy centers (Methods). Fig. 1 maps the entropy and inward flux of gas in a thin slice centered on one typical galaxy of $M_{\rm v} = 10^{12} M_\odot$ at $z = 2.5$. It demonstrates that the shock-heated, high-entropy, low-flux medium that fills most of the halo out to $R_{\rm v}$ and beyond is penetrated by three massive, narrow streams of low-entropy gas. They flow into the central disc with a radial flux per solid angle several times the average virial flux as estimated from eq. (1). The opening angle of a typical stream at $R_{\rm v}$ is $20^\circ - 30^\circ$, so the streams cover a total area of $\sim 0.4\,{\rm rad}^2$, namely a few percent of the sphere. The flux map in Fig. 2 displays the three-dimensional structure of the streams and uncovers clumps along two of them.

The penetration through the halo into the center can be evaluated via the profiles of total mass inflow rate $\dot{M}(r)$ through shells of radius r, obtained by integrating $\rho\, v_r r^2 d\Omega$ over a shell, where ρ is the gas density, v_r the radial velocity, and $d\Omega$ the angular-area element. Fig. 3 displays the average flux profile over the simulated galaxies of $M_{\rm v} \simeq 10^{12} M_\odot$ at $z = 2.5$, on top of the profiles of four representative cases (Supplementary Information).

The average profile reveals that the total flux remains roughly constant from well outside the virial radius ($\sim$90 kpc) all the way to the disc vicinity inside $r \sim 15$ kpc. The inflow rate of $\sim 100\, M_\odot\, \mathrm{yr}^{-1}$ is consistent with the virial growth rate predicted by eq. (1). Apparently, the flux decay while traveling through the halo is roughly compensated by the higher cosmological flux when that gas entered the halo, eq. (1), leading to the apparent constancy of flux with radius.

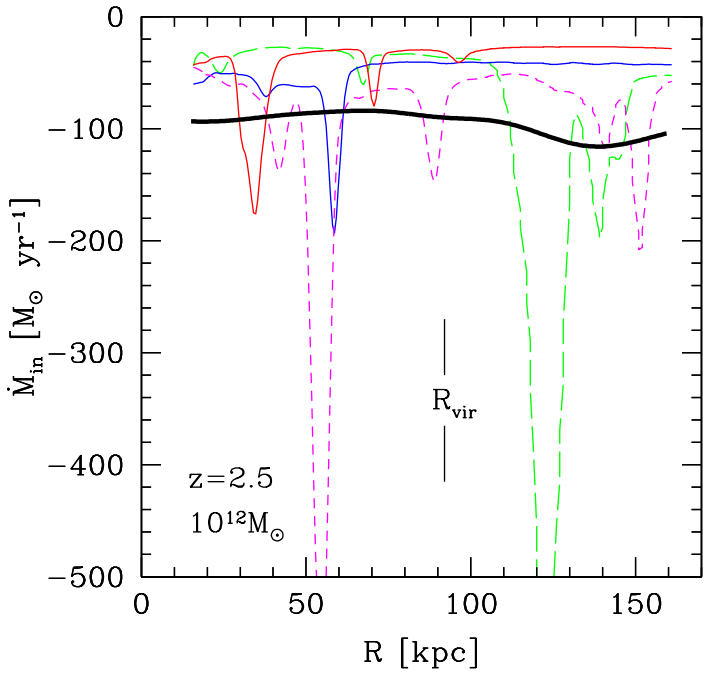

Figure 3. Gas inflow rate through spherical shells of radius r, from the disc vicinity to almost twice the halo virial radius. The thick black curve is the average over the simulated galaxies of $M_{\rm v} \simeq 10^{12} M_\odot$ at $z = 2.5$, showing deep penetration at a roughly constant rate. The colored curves refer to four representative galaxies, two showing clumps of $\mu \gtrsim 0.1$ and two with smoother flows involving only mini-minor clumps of $\mu < 0.1$.

Abundance of Gas Inflow Rates

In order to relate to the observed abundance of galaxies as a function of SFR, we use the MareNostrum flux profiles to evaluate $n(>\dot{M})$, the comoving number density of galaxies with an instantaneous gas-flow rate higher than $\dot{M}$. We first extract from the simulated flux profiles the conditional probability distribution at a given mass, $P(\dot{M}|M_{\rm v})$. It is done by sampling the $\dot{M}(r)$ profiles uniformly in r, given that the inflow velocity along the stream is roughly constant (Methods & Supplementary Informations). This is convolved with the halo mass function $n(M_{\rm v})$, based on the Sheth-Tormen approximation,[21] to provide

$$n(\dot{M}) = \int_0^\infty P(\dot{M}|M_{\rm v})\, n(M_{\rm v})\, dM_{\rm v} \ . \qquad (2)$$

The desired cumulative abundance $n(>\dot{M})$ is obtained by integration from $\dot{M}$ to infinity, and shown at $z = 2.2$ in the upper curve of Fig. 4. We see that galaxies with $\dot{M} >$

$150\,M_\odot\,\mathrm{yr}^{-1}$ are expected at a comoving number density $n\sim 3\times10^{-4}\,\mathrm{Mpc}^{-3}$. Fluxes as high as $\dot{M}>500\,M_\odot\,\mathrm{yr}^{-1}$ are anticipated at $n\sim6\times10^{-5}\,\mathrm{Mpc}^{-3}$. It is encouraging to note that this theoretical prediction lies safely above the observed values as indicated by the symbols. However, the difference between the gas supply rate and SFR is only by a factor of order 2, confirming our earlier conclusion that once the gas reaches the disc, it should very efficiently convert into stars on a dynamical time scale.

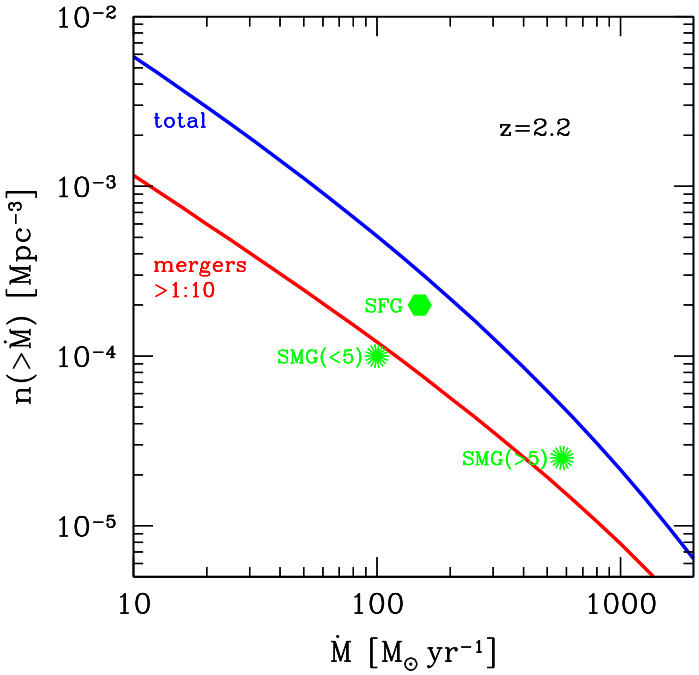

Figure 4. Comoving number density of galaxies with gas inflow rate higher than $\dot{M}$ at $z=2.2$, as predicted from our analysis of the cosmological simulation. The upper curve refers to total inflow, and the lower curve is limited to gas input by $\mu>0.1$ mergers. The symbols represent the vicinity of where the observed massive star-forming galaxies can be put once the observed SFR is identified with $\dot{M}$. The sBzK/BX/BM galaxies are marked SFG.[10] The SMGs brighter and fainter than 5 mJy are marked accordingly.[10,9] The gas inflow rate is sufficient for the SFR, but the small margin implies that the SFR must closely follow the gas supply. Most of the massive star formers should be observed while being fed by smooth flows rather than undergoing mergers.

Smooth Flows versus Mergers

By analyzing the clumpiness of the gas streams, we address the role of major mergers versus smooth flows in the disc buildup and star formation. Fig. 2 exhibits five obvious clumps along the streams. The sample profiles in Fig. 3 reveal the clumps as narrow peaks of inward flux. We evaluate each clump mass by integrating $M_{\rm clump}=\int\dot{M}(r)\,dr/v_r(r)$ across the peak, and estimate a mass ratio for the expected merger of the clump with the central galaxy by $\mu=M_{\rm clump}/(f_bM_{\rm v})$, ignoring mass loss in the clump on its way in. We term "merger" any major or minor merger of $\mu\geq0.1$, as opposed to "mini-minor" mergers where $\mu<0.1$, which we refer to as "smooth" flows. Fig. 3 shows two galaxies with $\mu>0.1$

clumps inside $2R_{\rm v}$, and two involving only smoother flows in this radius range. We find that about half the mass is flowing in as mergers and half as smoother flows. However, for $M_{\rm v}\sim10^{12}M_\odot$ and $z\sim2.2$, the duty cycle of mergers in a galaxy history is only $\eta\lesssim0.1$, i.e., less than one out of ten galaxies is undergoing a merger at any given time.

It is encouraging to note that a similar estimate can be obtained from an EPS-based prediction, by reading from Fig. 6 of Neistein & Dekel[7] the rate $dN/d\omega$ of mergers into a halo $M_{\rm v}$ with mass ratio $>\mu$. For the typical starburst duration we assume $\Delta t\simeq0.1R_{\rm v}/V_{\rm v}$ (namely $\sim50\,$Myr at $z=2.5$), based on merger simulations.[22] A similar estimate is derived from the typical peak width in the $\dot{M}(r)$ profiles of Fig. 3, given streaming at the virial velocity $(GM_{\rm v}/R_{\rm v})^{1/2}\sim220\,{\rm km\,s^{-1}}$. This leads to $\eta=(dN/d\omega)\Delta t\simeq0.09$ for $M=2\times10^{12}M_\odot$ at $z=2.2$.

The lower curve in Fig. 4 is obtained similarly to the upper curve, but using only the gas fed by mergers ($\mu>0.1$). From the difference between the two curves we learn that among the galaxies of a given $\dot{M}$, only about a quarter are predicted to be seen while undergoing a merger. The fact that the symbol representing typical SFGs (sBzK and BM/BX galaxies) lies well above the lower, merger curve indicates that in most of them the star formation is driven by smooth flows rather than mergers. This may explain why these galaxies maintain an extended thick disc while doubling their mass over a halo crossing time.[4] On the other hand, as seen in Fig. 4, about half the bright SMGs (with flux >5mJy), and most of the fainter SMGs, are consistent with being merger-induced starbursts.[10]

Conclusion

The cosmological gas accretion rate into galactic haloes of $M_{\rm v}\gtrsim10^{12}M_\odot$ at $z=$ 2-3 is sufficient to explain the observed abundance of high star-formation rates, but only barely so. This implies that the transition from accretion at the virial radius to SFR in a central galaxy has to be very efficient.[2,4] Indeed, our simulation reveals that the cold gas is poured into the center of each halo via a few steady narrow streams, riding the dark-matter filaments of the cosmic web, and very effectively penetrating through the shock-heated medium that otherwise fills the halo. A large fraction of the baryons comes in as mini-minor clumps, or smooth flows. The more clumpy input component, which involves about half the incoming mass, is also gas rich. We read from Fig. 4 that at a given SFR, the abundance of galaxies forming stars by smooth streams is ~3 times higher than the merger-induced starbursts. The Star-Forming Galaxies are predominantly fed by smooth flows, and can thus be interpreted as "Stream-Fed Galaxies". The same is true for a non-negligible fraction of the bright SMGs, the rest being driven by $\mu\gtrsim0.1$ mergers.

One should note that the intense SFR associated with streams must be limited to galaxies of $\sim10^{11}M_\odot$ in order not to violate the observational constraints on the overall density of SFR at these epochs.[11,23] Star formation must be suppressed in much smaller galaxies, possibly due to stellar feedback.[24,25,8] For completeness, it is also interesting to note that about half the $\sim10^{11}M_\odot$ galaxies at $z\sim2.2$ have suffered a major merger during the preceding ~1.5Gyr, as well as drastic heating by an expanding shock[20,26] (Libeskind,

N., Birnboim, Y. & Dekel, A., in preparation), so many of them are expected to be compact spheroids of low SFR rather than extended massive star formers.[27,28] This implies an even tighter correspondence between the gas input rate and the SFR.

METHODS

Halo Growth by EPS.

Neistein et al.[6] used the EPS[13] theory of cosmological clustering into spherical haloes in virial equilibrium to derive a robust approximation for the average growth rate of halo virial mass $M_{\rm v}$,

$$d\ln M_{\rm v}/d\omega = -(2/\pi)^{1/2}[\sigma^2(M_{\rm v}/q) - \sigma^2(M_{\rm v})]^{-1/2}, \quad \omega \equiv \delta_{\rm c}/D(t)\ . \tag{3}$$

The time variable ω, which makes the expression time invariant, is inversely proportional to $D(t)$, the linear growth rate of density fluctuations at time t in the assumed cosmology, with $\delta_{\rm c} \simeq 1.68$. The power spectrum of initial density fluctuations enters via the variance $\sigma^2(M_{\rm v})$. The constant q is 2.2 with an uncertainty of ± 0.1 intrinsic to the EPS theory. Eq. (3) has been confirmed to resemble the assembly rate in cosmological N-body simulations.[14]

The MareNostrum simulation.

The cosmological simulation[5,29] used in the present analysis has been performed with the Eulerian AMR code RAMSES[30] on 2,048 processors of the MareNostrum supercomputer. The code simulates the coupled gas and dark-matter dynamics, using a Particle-Mesh scheme for the dark-matter component and a second-order Godunov scheme for the gas component. In order to describe the formation of dense star-forming discs, the code includes metal-dependent radiative cooling, UV heating by a standard photo-ionizing background, star formation, supernovae feedback and metal enrichment. The simulation box of comoving $50\,h^{-1}$Mpc involved $1,024^3$ dark-matter particles and 4×10^9 gas cells. Using a quasi-Lagrangian refinement strategy, the spatial resolution reaches $\sim 1\,h^{-1}$kpc in physical units at all times. The dark-matter particle mass is $\sim 10^8 M_\odot$, so each of the haloes studied here consists of $\sim 10^4$ particles within the virial radius. Since one can reliably describe the formation of haloes down to ~ 100 particles,[31] namely $\sim 10^{10} M_\odot$, the $10^{12} M_\odot$ haloes addressed here are two orders of magnitude above the minimum halo mass. This simulation allows us to capture the important properties of gas accretion into galaxies in more than 100 haloes of $\sim 10^{12} M_\odot$ at $z \sim 2.5$, thus providing a large statistical sample. A first analysis of galaxies from this simulation[5] have confirmed the bi-modal nature of cold flows and hot media as a function of mass and redshift.[8]

Our current analysis is based on robust features that are properly simulated, such as the large-scale structure of the streams, the total flux in them, and the gas clumps more massive than $\sim 10^{10} M_\odot$. However, the finite resolution does not allow a fair treatment of small-scale gas phenomena such as turbulence in the hot gas, ram-pressure stripping of clumps, hydrodynamical instabilities at the stream boundaries, and the formation of small

clumps. Furthermore, the current resolution does not allow a detailed study of the discs that form at the halo centers as the disc thickness is barely resolved. More accurate analysis of the fine stream structure and disc buildup should await simulations of higher resolution.

Computing the flux abundance.

In order to evaluate the conditional probability $P(\dot{M}|M_{\rm v})$ for eq. (2), we measure $P_0(\dot{M}|M_0)$ from a fair sample of MareNostrum galaxies in haloes of $M_0 = 10^{12} M_\odot$ at $z_0 = 2.5$ (Supplementary Information). We then generalize it to other masses $M_{\rm v}$ using the scaling from eq. (1), $\dot{M} \propto M_{\rm v}^{1.15}$, namely

$$P(\dot{M}|M_{\rm v}) = P_0[\dot{M}(M_0/M_{\rm v})^{1.15}|M_0] \ . \qquad (4)$$

At $z \sim 2.5$, this estimate of $\dot{M}$ is good to within a factor of two for $M_{\rm v} \leq 10^{13} M_\odot$, beyond which it becomes a more severe overestimate (Goerdt et al., in preparation). The results for other redshifts ($z > 2$) are obtained using the scaling from eq. (1), $\dot{M} \propto (1+z)^{2.25}$.

Received ; Accepted .

---

1. Chapman, S. C., Smail, I., Blain, A. W. & Ivison, R. J. A Population of Hot, Dusty Ultraluminous Galaxies at $z \sim 2$. *Astrophys. J.* **614**, 671–678 (2004).

2. Genzel, R. *et al.* The rapid formation of a large rotating disk galaxy three billion years after the Big Bang. *Nature* **442**, 786–789 (2006).

3. Förster Schreiber, N. M. *et al.* SINFONI Integral Field Spectroscopy of $z \sim 2$ UV-selected Galaxies: Rotation Curves and Dynamical Evolution. *Astrophys. J.* **645**, 1062–1075 (2006).

4. Genzel, R. *et al.* From rings to bulges: evidence for rapid secular galaxy evolution at $z \sim 2$ from integral field spectroscopy in the SINS survey. ArXiv e-prints 0807.1184 (2008).

5. Ocvirk, P., Pichon, C. & Teyssier, R. Bimodal gas accretion in the MareNostrum galaxy formation simulation. ArXiv e-prints 0803.4506 (2008).

6. Neistein, E., van den Bosch, F. C. & Dekel, A. Natural downsizing in hierarchical galaxy formation. *Mon. Not. R. Astron. Soc.* **372**, 933–948 (2006).

7. Neistein, E. & Dekel, A. Merger Rates of Dark-Matter Haloes. ArXiv e-prints 0802.0198 (2008).

8. Dekel, A. & Birnboim, Y. Galaxy bimodality due to cold flows and shock heating. *Mon. Not. R. Astron. Soc.* **368**, 2–20 (2006).

9. Wall, J. V., Pope, A. & Scott, D. The evolution of submillimetre galaxies: two populations and a redshift cut-off. *Mon. Not. R. Astron. Soc.* **383**, 435–444 (2008).

10. Tacconi, L. J. *et al.* Submillimeter Galaxies at $z \sim 2$: Evidence for Major Mergers and Constraints on Lifetimes, IMF, and CO-H_2 Conversion Factor. *Astrophys. J.* **680**, 246–262 (2008).

11. Madau, P., Pozzetti, L. & Dickinson, M. The Star Formation History of Field Galaxies. *Astrophys. J.* **498**, 106–114 (1998).

12. Bouché, N. *et al.* Dynamical Properties of $z \sim 2$ Star-forming Galaxies and a Universal Star Formation Relation. *Astrophys. J.* **671**, 303–309 (2007).

13. Lacey, C. & Cole, S. Merger rates in hierarchical models of galaxy formation. *Mon. Not. R. Astron. Soc.* **262**, 627–649 (1993).

14. Neistein, E. & Dekel, A. Constructing merger trees that mimic N-body simulations. *Mon. Not. R. Astron. Soc.* **383**, 615–626 (2008).

15. Genel, S. *et al.* Mergers and Mass Accretion Rates in Galaxy Assembly: The Millennium Simulation Compared to Observations of $z \sim 2$ Galaxies. (2008).

16. Komatsu, E. *et al.* Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ArXiv e-prints 0803.0547 (2008).

17. Birnboim, Y. & Dekel, A. Virial shocks in galactic haloes? *Mon. Not. R. Astron. Soc.* **345**, 349–364 (2003).

18. Binney, J. On the origin of the galaxy luminosity function. *Mon. Not. R. Astron. Soc.* **347**, 1093–1096 (2004).

19. Kereš, D., Katz, N., Weinberg, D. H. & Davé, R. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).

20. Birnboim, Y., Dekel, A. & Neistein, E. Bursting and quenching in massive galaxies without major mergers or AGNs. *Mon. Not. R. Astron. Soc.* **380**, 339–352 (2007).

21. Sheth, R. K. & Tormen, G. An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *Mon. Not. R. Astron. Soc.* **329**, 61–75 (2002).

22. Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R. & Dekel, A. The effect of galaxy mass ratio on merger-driven starbursts. *Mon. Not. R. Astron. Soc.* **384**, 386–409 (2008).

23. Hopkins, A. M. On the Evolution of Star-forming Galaxies. *Astrophys. J.* **615**, 209–221 (2004).

24. Dekel, A. & Silk, J. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *Astrophys. J.* **303**, 39–55 (1986).

25. Dekel, A. & Woo, J. Feedback and the fundamental line of low-luminosity low-surface-brightness/dwarf galaxies. *Mon. Not. R. Astron. Soc.* **344**, 1131–1144 (2003).

26. Dekel, A. & Birnboim, Y. Gravitational quenching in massive galaxies and clusters by clumpy accretion. *Mon. Not. R. Astron. Soc.* **383**, 119–138 (2008).

27. Kriek, M. *et al.* Spectroscopic Identification of Massive Galaxies at $z \sim 2.3$ with Strongly Suppressed Star Formation. *Astrophys. J. Lett.* **649**, L71–L74 (2006).

28. van Dokkum, P. G. *et al.* Confirmation of the Remarkable Compactness of Massive Quiescent Galaxies at $z \sim 2.3$: Early-Type Galaxies Did not Form in a Simple Monolithic Collapse. *Astrophys. J. Lett.* **677**, L5–L8 (2008).

29. Prunet, S. *et al.* Initial Conditions for Large Cosmological Simulations. ArXiv e-prints 0804.3536 (2008).

30. Teyssier, R. Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. *Astron. Astrophys.* **385**, 337–364 (2002).

31. Rasera, Y. & Teyssier, R. The history of the baryon budget. Cosmic logistics in a hierarchical universe. *Astron. Astrophys.* **445**, 1–27 (2006).

32. Press, W. H. & Schechter, P. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophys. J.* **187**, 425–438 (1974).

33. Springel, V. *et al.* Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).

34. Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B. & Blaizot, J. Modelling the galaxy bimodality: shutdown above a critical halo mass. *Mon. Not. R. Astron. Soc.* **370**, 1651–1665 (2006).

35. Croton, D. J. *et al.* The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *Mon. Not. R. Astron. Soc.* **365**, 11–28 (2006).

36. Bower, R. G. *et al.* Breaking the hierarchy of galaxy formation. *Mon. Not. R. Astron. Soc.* **370**, 645–655 (2006).

37. Cattaneo, A., Dekel, A., Faber, S. M. & Guiderdoni, B. *et al.* Downsizing by Shutdown in Red Galaxies. ArXiv e-prints 0801.1673 (2008).

Acknowledgments We acknowledge stimulating discussions with N. Bouche, R. Genzel, A. Kravtsov, A. Pope, A. Sternberg & J. Wall. This research has been supported by the France-Israel Teamwork in Sciences, the German-Israel Science Foundation, NASA ATP, and a Minerva fellowship (TG). We thank the computer resources and technical support by the Barcelona Centro Nacional de Supercomputacion. The simulation is part of the Horizon collaboration.

Author Information Correspondence and requests for materials should be addressed to A.D. (dekel@phys.huji.ac.il).

SUPPLEMENTARY INFORMATION

This is an extension of the Letter, aimed at providing further details, in support of the results reported in the main body of the Letter.

1 On the origin of narrow streams

Dekel & Birnboim 2006[8] (hereafter DB06) put forward the conjecture that at redshifts higher than $z_{\rm crit} \sim 2$, narrow cold streams penetrate deep into the dark-matter haloes even when the haloes are more massive than the shock-heating scale, $M_{\rm shock} \lesssim 10^{12} M_\odot$. This prediction is summarized in Fig. 5.

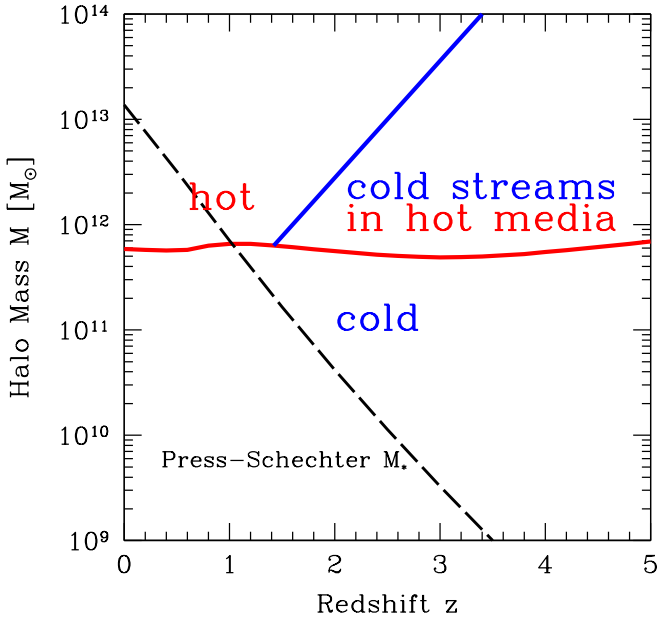

Figure 5. Analytic prediction for the regimes dominated by cold flows and shock-heated medium in the plane of halo mass and redshift, based on Fig. 7 of DB06. The nearly horizontal curve marks the robust threshold mass for a stable shock based on spherical infall analysis, $M_{\rm shock}(z)$. Below this curve the flows are predicted to be predominantly cold and above it a shock-heated medium is expected to extend out to the halo virial radius. The inclined solid curve is the conjectured upper limit for cold streams, valid at redshifts higher than $z_{\rm crit} \sim 2$. The hot medium in haloes of $M_{\rm v} > M_{\rm shock}$ at $z > z_{\rm crit}$ is predicted to host penetrating cold streams, while haloes of a similar mass at $z < z_{\rm crit}$ are expected to be all hot, shutting off most of the gas supply to the inner galaxy. Also shown is the characteristic Press-Schechter halo mass $M_*(z)$; it is much smaller than $M_{\rm shock}$ at $z > 2$.

The critical condition for a stable virial shock is that the radiative cooling rate behind the shock is slower than the compression rate, $t_{\rm cool}^{-1} < t_{\rm comp}^{-1}$, allowing the buildup of pressure

support behind the shock against global gravitational collapse. Based on a spherical analysis, DB06 found that a virial shock should exist in dark-matter haloes above a threshold mass $M_{\rm shock} \lesssim 10^{12} M_\odot$ that is rather constant in time, at an actual value that is sensitive to the metallicity of the gas in the halo. The existence of such a threshold mass and its value as a function of redshift have been confirmed by the analysis of cosmological simulations.[19,8,20,5] However, at high redshifts, even above the threshold mass, a shock is not expected to develop along narrow, cold, radial streams that penetrate through the halo, because the cooling there is more efficient than in the surrounding halo.

The appearance of intense streams at high z, as opposed to their absence at low z, reflects the interplay between the shock-heating scale and the independent characteristic scale of nonlinear clustering, i.e., the Press-Schechter[32] mass M_* that corresponds to the typical dark-matter haloes forming at a given epoch. The key difference between the two epochs is that the rapid growth of M_* with time, as seen in Fig. 5, makes $M_{\rm shock} \gg M_*$ at $z > 2$ while $M_{\rm shock} \sim M_*$ at lower redshifts.

The following argument explains why rare dark-matter haloes of $M_{\rm v} \gg M_*$ are fed by narrow streams, while typical haloes $\sim M_*$ accrete from a wide angle. The large-scale structure of dark matter is expected to be roughly self-similar in time when masses are measured in terms of M_* and densities in terms of the background universal density. We learn from cosmological N-body simulations[19,33] that this is indeed the case for the large-scale filamentary structure: the characteristic width of the filaments is comparable to $R_* \propto M_*^{1/3}$, the typical size of an M_* halo, while the typical filament length is larger by about an order of magnitude and scales similarly in time. This means that while the infall pattern into $\sim M_*$ haloes is practically a wide-angle, nearly spherical pattern, the infall into $M_{\rm v} \gg M_*$ haloes is along a few well-defined filaments that are thin relative to the halo size. Assuming that at any given epoch the accretion rate of dark matter, $\dot{M}$, is roughly proportional to the halo mass $M_{\rm v}$ (Eq. 1 of the Letter), while the virial densities in haloes of all masses are the same (by definition), this geometrical difference implies that the densities in the filaments penetrating $M_{\rm v} \gg M_*$ haloes are higher by a factor of a few than the typical densities in their host haloes. This is demonstrated in Fig. 6.

Assuming that the density of the gas flowing along the filaments scales with the dark-matter density, and that the infall velocity is comparable to the halo virial velocity, we conclude that the cooling rate in the filaments feeding an $M_{\rm v} \gg M_*$ halo should be higher by a factor of a few than in the surrounding spherical halo. If the compression rate in the filaments is comparable to that in the host halo, this implies that the thin filaments have a harder time supporting a stable shock. As a result, the critical halo mass for shock heating in the filaments feeding it must be larger by a factor of a few. This is the case for $M_{\rm v} \gtrsim M_{\rm shock}$ haloes at high redshifts.

A crude estimate led DB06 to the conjectured upper limit for penetrating streams shown in Fig. 5:

$$M_{\rm stream} \sim \frac{M_{\rm shock}}{f M_*} M_{\rm shock} \quad \text{for} \quad f M_* < M_{\rm shock} \ , \tag{5}$$

where the characteristic width of the streams is $\propto (f M_*)^{1/3}$, with f a factor of order a few.

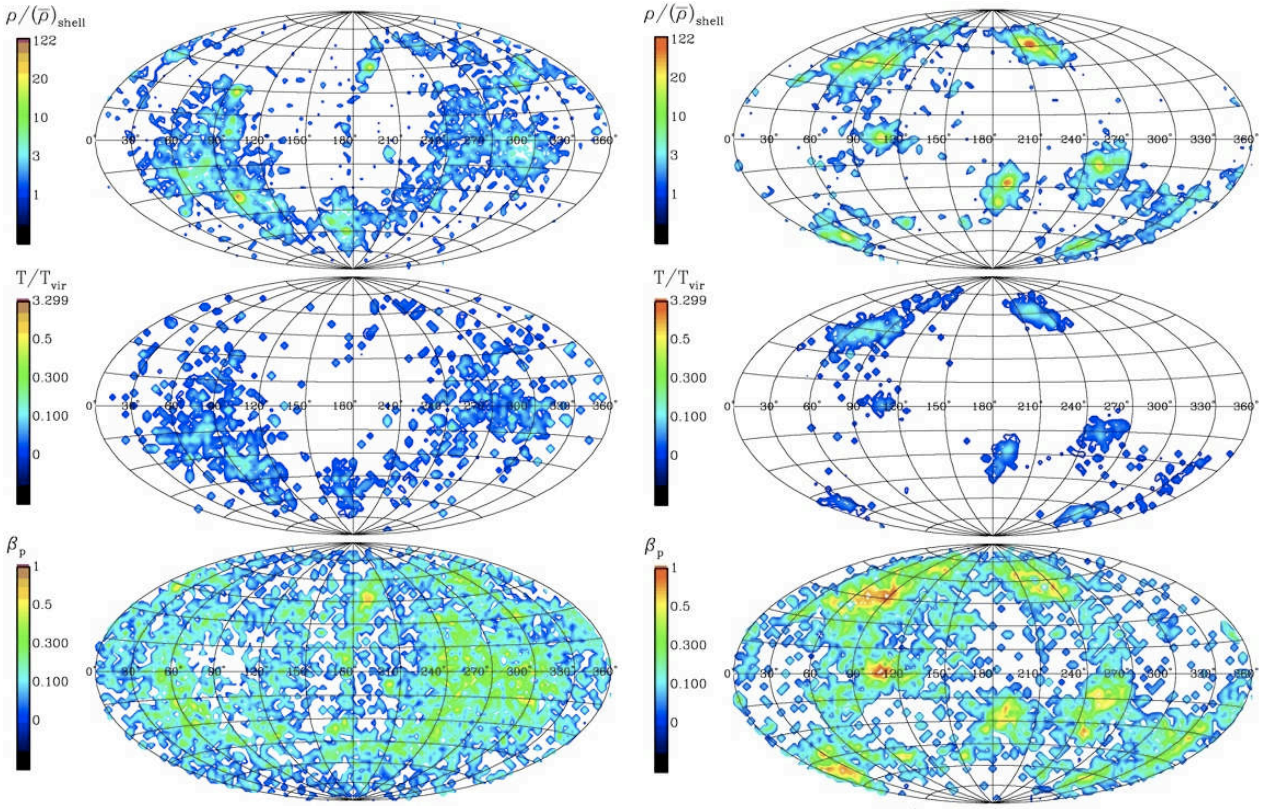

Figure 6. The pattern of dark-matter inflow in a shell $(1-3)R_{\rm v}$ outside two haloes from a cosmological N-body simulation at $z=0$ (based on P. Seleson & A. Dekel, in preparation). **Left:** a typical halo with $M_{\rm v}\sim M_*$. **Right:** a rare halo with $M_{\rm v}\gg M_*$. In terms of the different ways by which these two haloes are fed by dark-matter, they correspond to two haloes of the same mass $M_{\rm v}\sim 10^{12}M_\odot$, but at $z\sim 0$ and $z\sim 2-3$ respectively. Top to bottom: density, temperature and infall velocity. We see that the typical halo resides inside a broad filament so it is practically fed by wide-angle diffuse accretion. On the other hand, the rare halo is fed by a few narrow, dense, in-flowing filaments.

At low z, where $fM_*>M_{\rm shock}$, cold flows exist only for $M_{\rm v}<M_{\rm shock}$. At high z, where $fM_*<M_{\rm shock}$, cold streams appear even in $M_{\rm v}>M_{\rm shock}$ haloes where shocks have heated part of the gas, as long as $M_{\rm v}<M_{\rm stream}$. The critical redshift $z_{\rm crit}$ separating these two regimes is defined by

$$fM_*(z_{\rm crit})=M_{\rm shock}\ . \qquad (6)$$

This crude maximum mass for cold streams is shown in Fig. 5 for an arbitrary choice of $f=3$.

A preliminary analysis of the MareNostrum simulation[5] confirms the prediction of eq. (5), when taking into account the lower metallicities in the simulation compared to that assumed in the analytic calculation. The streams analyzed in the current Letter, in dark-matter haloes of $M_{\rm v}=10^{12}M_\odot$ at $z=2.5$, represent an encouraging confirmation of the validity of the DB06 conjecture. Further analysis in progress (T. Goerdt et al., in

preparation) indicates, for example, that at $z = 2.5$, the fraction of inflow in cold streams drops by a factor of three at $M_{\rm v} \simeq 2 \times 10^{13} M_\odot$, much in the spirit of the crude prediction of Fig. 5. The permitted cold gas supply by streams in massive haloes at high redshift, followed by a shutdown above $M_{\rm shock}$ at low redshifts, turn out to provide good match to many observed galaxy properties when these features are incorporated in semi-analytic simulations of galaxy formation.[34–37]

2 Maps of entropy, flux and density for several galaxies

Figs. 7 to 9 extend the visual information provided by Figs. 1 and 2 of the Letter. They display different gas properties that highlight the structure and kinematics of the cold streams in three simulated galaxies of $M_{\rm v} = 10^{12} M_\odot$ at $z = 2.5$.

The entropy maps show $\log(T/\rho^{2/3})$ where the temperature and gas density are in units of the virial temperature and mean density within the halo virial radius $R_{\rm v}$. They exhibit the virial shock, covering most of the area of the virial sphere and sometimes extending beyond $2R_{\rm v}$. The narrow streams are of much lower entropy, by more than three orders of magnitude, comparable to the low entropy in the central disc they lead to. The boundaries between the streams and the hot medium within the virial radius are sharp and well defined. We also note that semi-cylindrical shocks sometimes partly surround the elongated streams long before they enter the halo virial radius.

The arrows mark the velocity field projected on the slice plane, and the flux color maps show the flow rate per solid angle, $\dot{m} = r^2 \rho \, v_r$. The flux inward is almost exclusively channeled through the narrow streams. This flux is several times the average over a sphere, $\dot{m}_{\rm vir} \simeq 8 \, M_\odot \, {\rm yr}^{-1} {\rm rad}^{-2}$. The opening angle of a typical stream at $R_{\rm v}$ is $20 - 30^\circ$, so the streams cover a total area of $\sim 0.4 \, {\rm rad}^2$, namely a few percent of the sphere. The velocity field in the hot medium is turbulent and sometimes showing vast outflows. The inward flux over most of the sphere area is negligible, both inside and outside the virial radius or the virial shock. The streaming velocities are supersonic, with a Mach number of order a few.

Although the streams tend to be rather radial when viewed on scales comparable to the halo virial radius, some of them flow in with impact parameters on the order of 10 kpc, comparable to the disc sizes. The steady high flux along a line of a rather fixed orientation with a non-negligible impact parameter is the source of angular momentum required for the buildup of an extended rotating disc (A. Zinger et al., in preparation).

The gas density maps, in units of the mean gas density within the virial radius, emphasize the narrowness of the streams, and reveal that they are typically denser than the surrounding medium by more than an order of magnitude. This confirms the prediction described in §1, and explains why a virial shock is avoided along the streams, allowing them to penetrate cold and unperturbed into the inner halo.

The column-density maps of the inflowing material are obtained by summing up the densities in grid cells along each line of sight inside the box of side 320 kpc. The cells that enter this sum are only those where the inward flux per solid angle is at least twice the

average over a sphere based on Eq. 1 of the Letter. The densities here are in units of cm^{-3}, after dividing by $0.6m_{\rm p}$. These maps highlight the three-dimensional configuration of radial streams, and the clumps along some of them.

Fig. 10 displays three-dimensional TIPSY* pictures of the radial influx $\dot{m}$, similar to Fig. 2 of the Letter. It shows the overall structure of the inflowing streams in 3D perspective for four simulated haloes. The pictures reveal that the typical configuration is of three major narrow streams. Some of the streams are straight lines, and others are curved. Some of the streams are of rather fixed width from well outside $R_{\rm v}$, and others display a conical shape, starting broad at large radii and getting narrower as they penetrate into the halo. The gas streams show dense clumps, with about half the stream mass in clumps of mass ratio $\mu > 0.1$, namely mass above $\sim 10^{10} M_\odot$. The rest is in smaller clumps, some clearly hidden below the resolution limit. Since these mini-minor clumps are not expected to cause significant damage to the central disc,[22] we can refer to them in this respect as "smooth" flows. It is not clear at this point to what extent the smooth component is truly smooth or built by mini-minor clumps, and whether the perfect smoothness has a physical origin or is merely a numerical artifact, but this distinction does not make a qualitative difference to our present discussion.

*http://www-hpcc.astro.washington.edu/tools/tipsy/tipsy.html

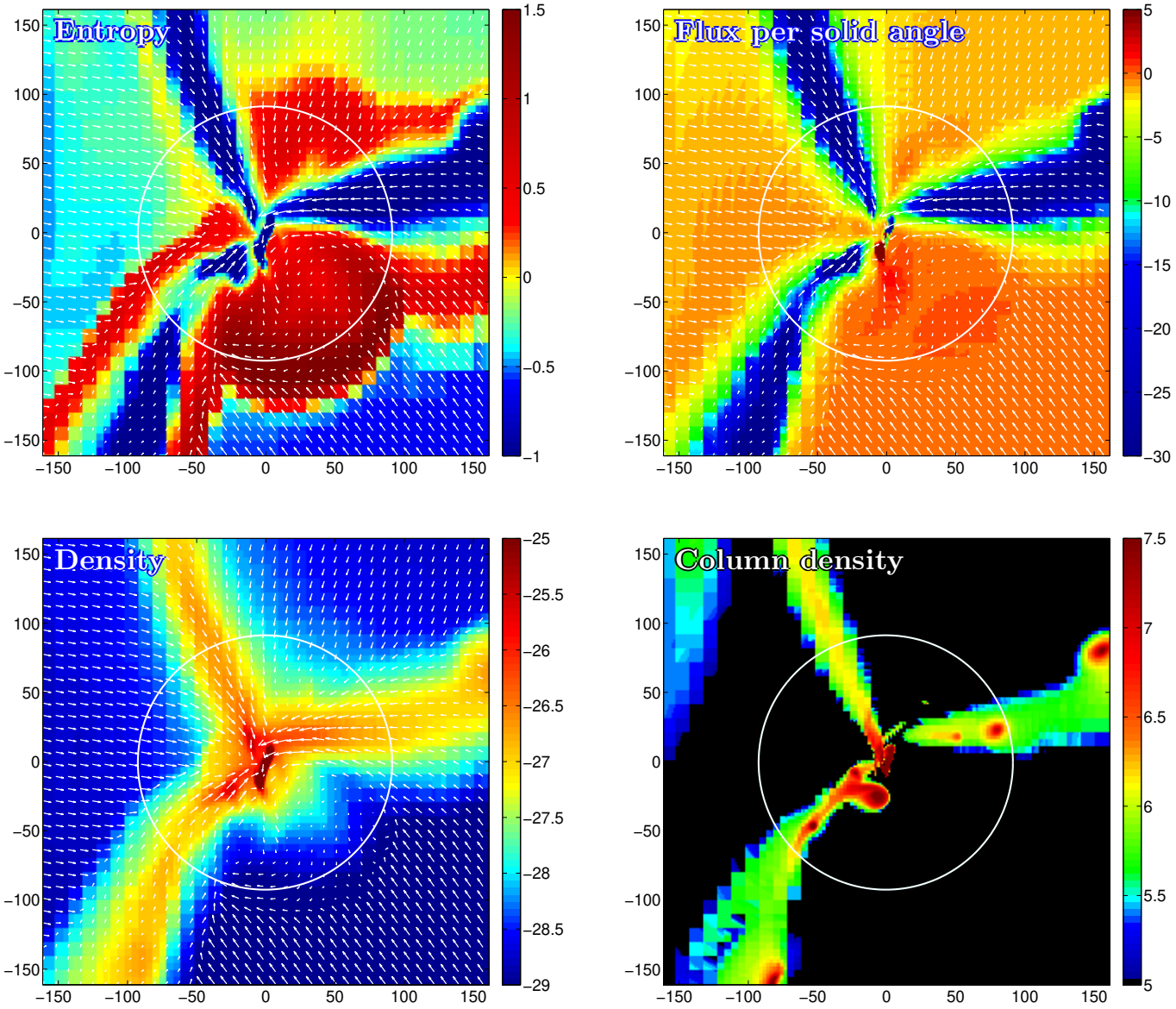

Figure 7. Gas in halo 314 of the MareNostrum simulation. Three maps refer to a thin equatorial slice. They show (a) entropy $\log K = \log(T/\rho^{2/3})$ in units of the virial quantities, (b) radial flux $\dot{m} = r^2 \rho\, v_r$ in $M_\odot\,\mathrm{yr}^{-1}\,\mathrm{rad}^{-2}$, and (c) log density in units of the mean gas density within R_v. The fourth panel shows log column density through the 3D box of side 320 kpc, considering only the cells where the radial flux inward is at least twice as high as the average over a shell based on Eq. 1 of the Letter. The circle marks the virial radius.

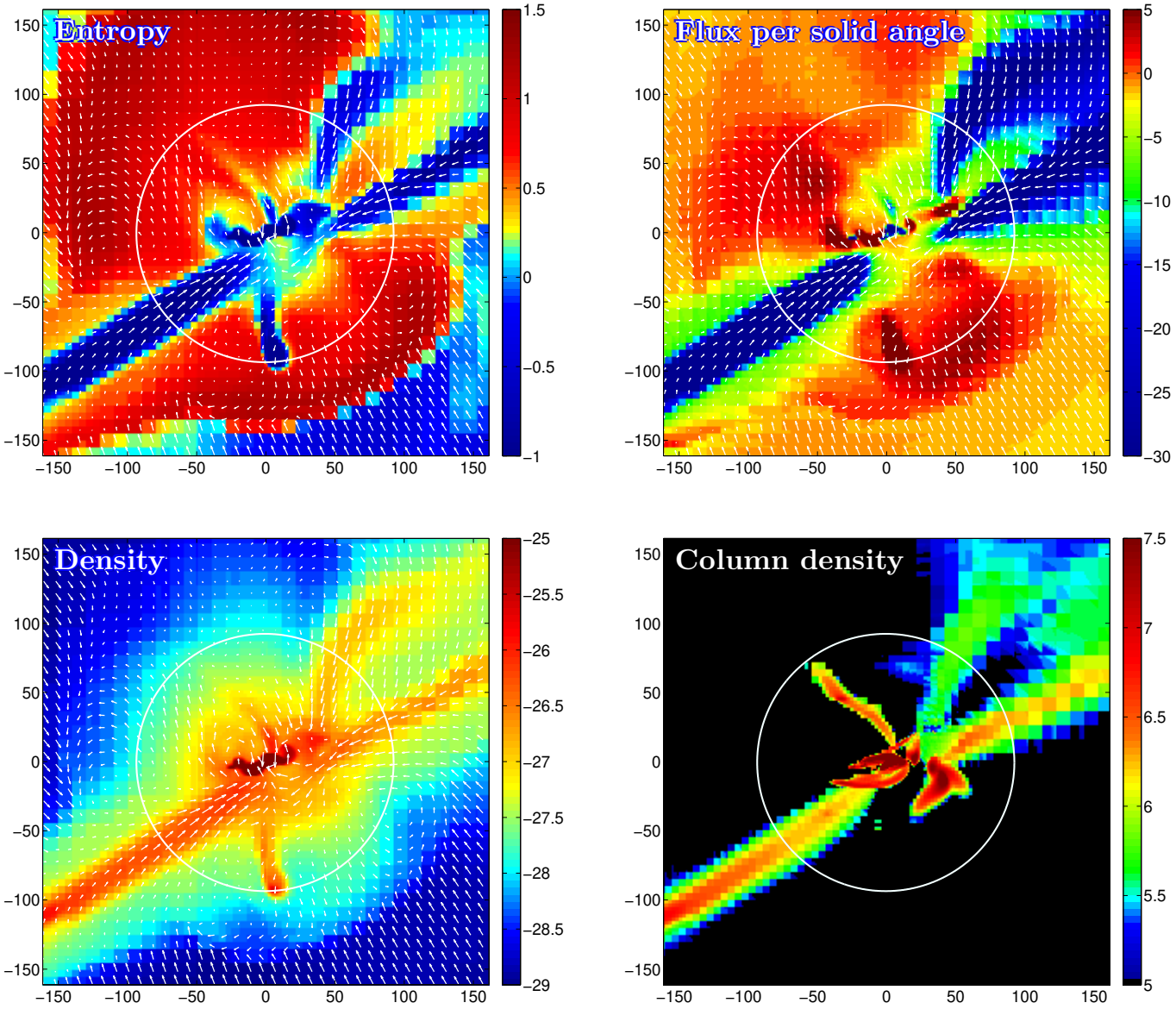

Figure 8. Gas in halo 303 of the MareNostrum simulation. Three maps refer to a thin equatorial slice. They show (a) entropy $\log K = \log(T/\rho^{2/3})$ in units of the virial quantities, (b) radial flux $\dot{m} = r^2 \rho\, v_r$ in $M_\odot\,\mathrm{yr}^{-1}\,\mathrm{rad}^{-2}$, and (c) log density in units of the mean gas density within R_v. The fourth panel shows log column density through the 3D box of side 320 kpc, considering only the cells where the radial flux inward is at least twice as high as the average over a shell based on Eq. 1 of the Letter. The circle marks the virial radius.

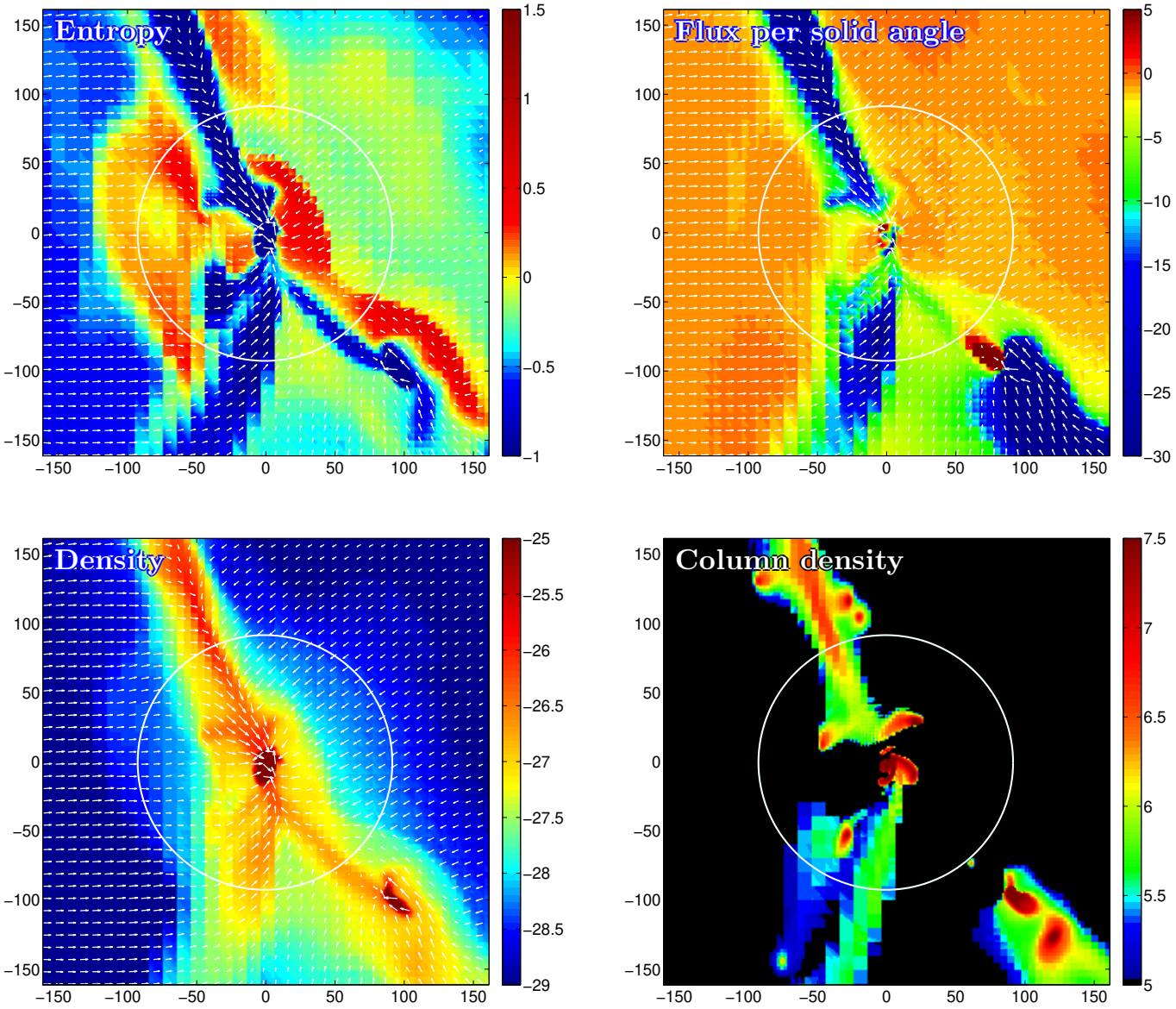

Figure 9. Gas in halo 311 of the MareNostrum simulation: Three maps refer to a thin equatorial slice. They show (a) entropy $\log K = \log(T/\rho^{2/3})$ in units of the virial quantities, (b) radial flux $\dot{m} = r^2 \rho\, v_r$ in $M_\odot\,\mathrm{yr}^{-1}\,\mathrm{rad}^{-2}$, and (c) log density in units of the mean gas density within R_v. The fourth panel shows log column density through the 3D box of side 320 kpc, considering only the cells where the radial flux inward is at least twice as high as the average over a shell based on Eq. 1 of the Letter. The circle marks the virial radius.

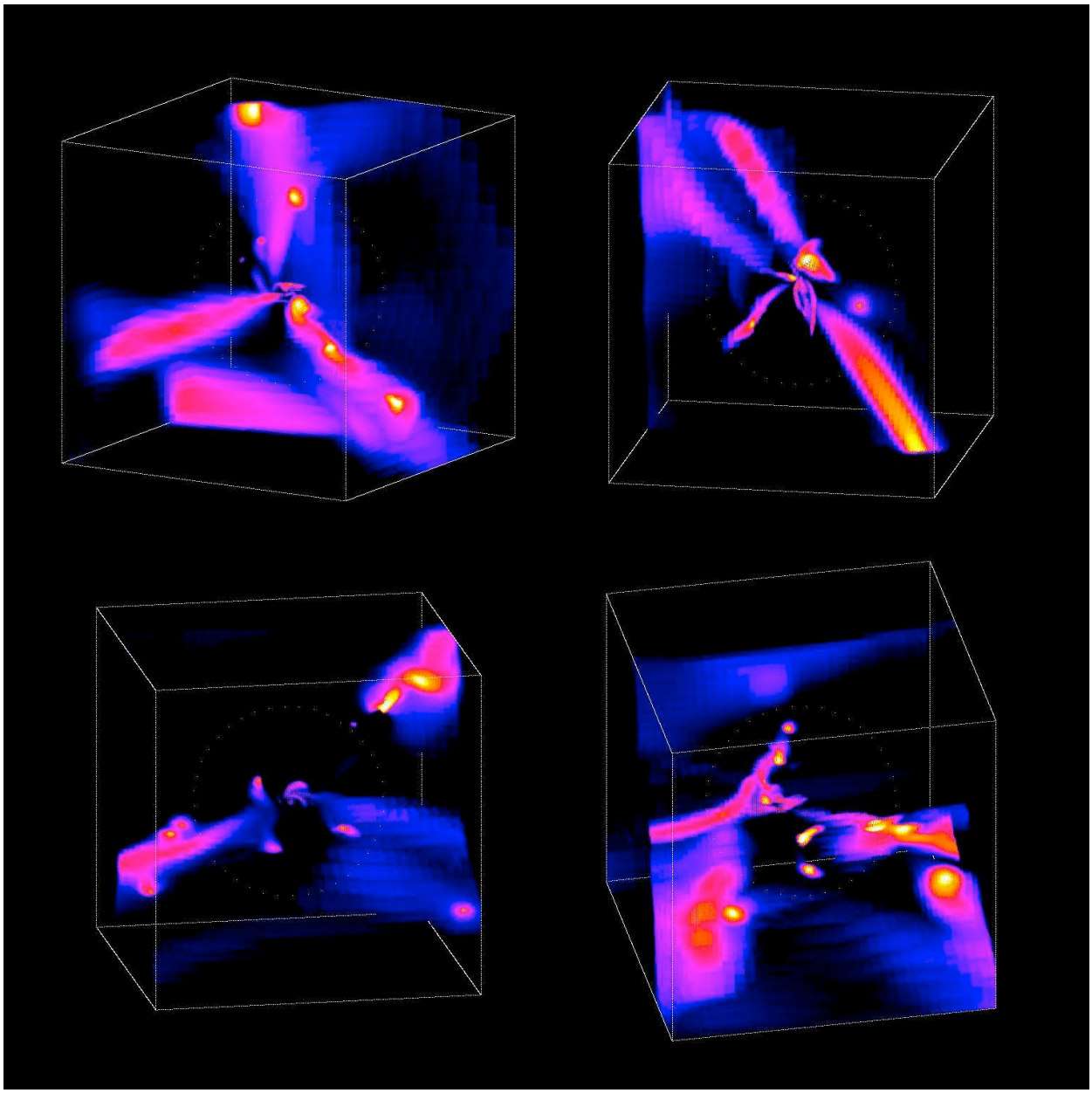

Figure 10. Inward flux in the three-dimensional boxes of side 320 kpc centered on galaxies 314, 303, 311 and 310 from the MareNostrum simulations. The colors refer to inflow rate per solid angle of point-like tracers at the centers of the cubic grid cells. The dotted circle marks the virial radius. All haloes show high-flux streams, some smooth and some with embedded clumps. Galaxy 310 (bottom right) is undergoing multiple minor mergers due to the particularly clumpy streams.

3 Flux profiles and probability distribution

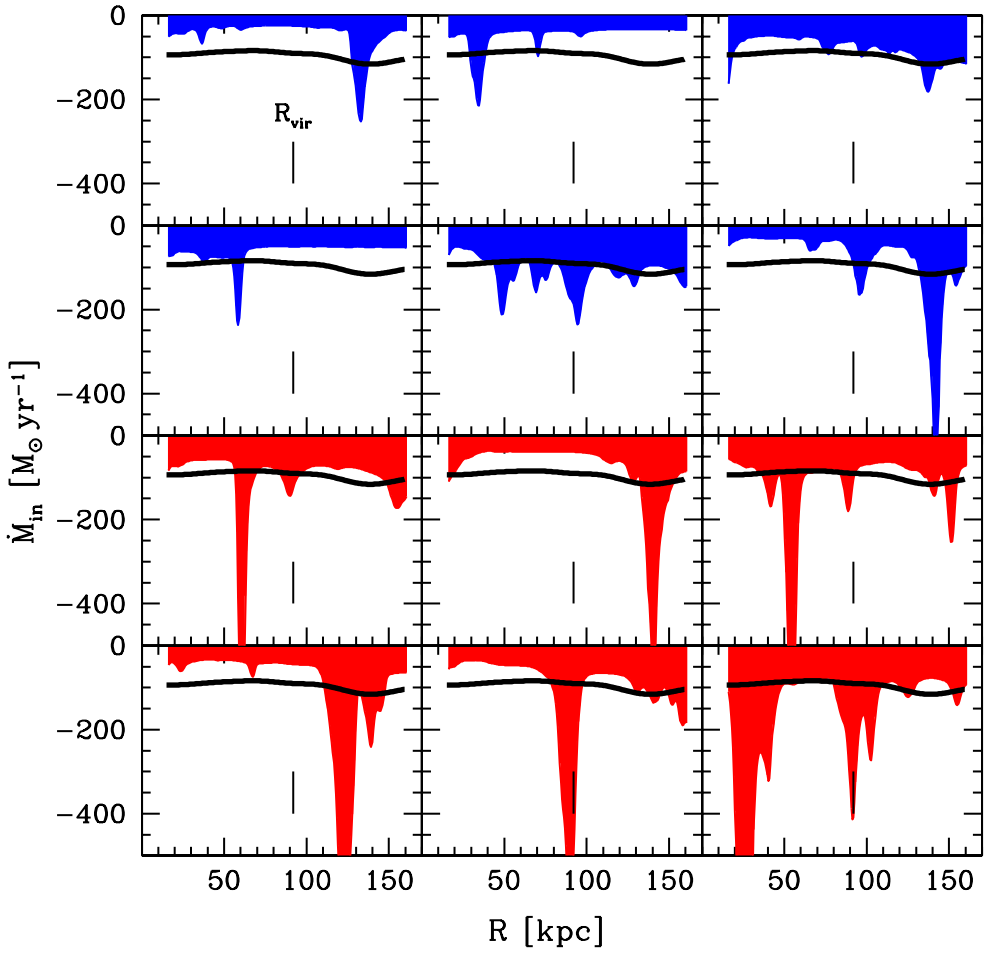

Figure 11. Profiles of total gas inflow rate through spherical shells as in Fig. 3 of the Letter. Shown here are twelve galaxies of $M_{\rm v} \simeq 10^{12} M_\odot$ at $z = 2.5$, randomly chosen from the simulation. The lower six panels show clumps that correspond to mergers of mass ratio $\mu > 0.1$, while the upper six are fed by smoother flows with only mini-minor mergers of $\mu < 0.1$.

Fig. 11 is an extension of Fig. 3 of the Letter, presenting the influx profiles of twelve galaxies, all with $M_{\rm v} \simeq 10^{12} M_\odot$ at $z = 2.5$, chosen at random from the MareNostrum simulation. The profiles extend from $r = 15\,$kpc, the disc vicinity, to $r = 160\,$kpc, almost twice the virial radius of $R_{\rm v} \simeq 90\,$kpc. Six galaxies turn out to show clumps leading to mergers of $\mu > 0.1$, and the rest show only smaller clumps in smoother flows. One can read from the relative width of the clumps in the figure that the duty cycle for $\mu > 0.1$ clumps in each individual galaxy is less than 0.1. By comparing the areas above the individual profiles with the average for galaxies of that mass and redshift, one can see that on average only about half the stream mass is in clumps.

Fig. 12 shows the probability distribution of $\dot{M}$ for the fiducial galaxies of $M_{\rm v} = M_0 = 10^{12} M_\odot$ at $z = 2.5$. It has been derived as explained in the Methods section of the Letter, by uniform sampling of the flux profiles shown in Fig. 11. This $P_0(\dot{M}|M_0)$ is used in Eq. 4

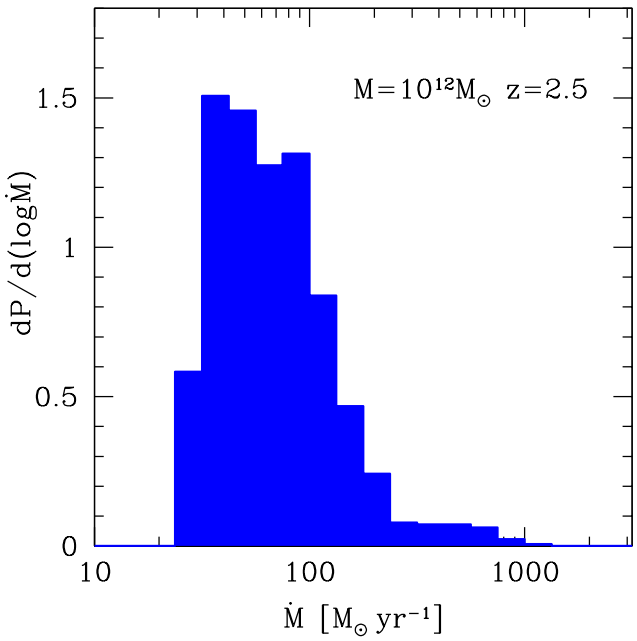

Figure 12. The conditional probability distribution $P(\dot{M}|M_{\rm v})$ for the fiducial case $M_{\rm v}=M_0=10^{12}M_\odot$ at $z=2.5$.

of the Letter to derive the conditional probability for other halo masses, $P(\dot{M}|M_{\rm v})$, which is then used in Eq. 2 to derive the abundance of galaxies with a given $\dot{M}$. The tail or the distribution at $\dot{M}>200\,M_\odot\,{\rm yr}^{-1}$ is dominated by $\mu>0.1$ mergers, while the main body of the distribution is mostly due to smoother streams. Recall that the average is about $100\,M_\odot\,{\rm yr}^{-1}$.

4 The abundance as a function of mass and redshift

As described in Methods, we derived the conditional probability distribution $P(\dot{M}|M_{\rm v})$ from the simulated haloes at $z=2.5$ for a fiducial halo mass $M_{\rm v}=10^{12}M_\odot$, shown in Fig. 12. We then used the scaling of eq. 1 of the Letter to obtain an estimate for other halo masses. Preliminary analysis of more massive haloes at that redshift (T. Goerdt et al., in preparation) indicates that the actual flux starts dropping below the adopted estimate in haloes more massive than $M_{\rm stream}\sim 10^{13}M_\odot$. For a first crude estimate of the effect this might have on our results shown in Fig. 4 of the Letter, we re-compute the comoving number density $n(>\dot{M})$ as described in the main text, but now limit the halo mass range that contributes to $\dot{M}$ by an upper cutoff at $M_{\rm stream}$. Fig. 13 shows the results for different values of $M_{\rm stream}$. We see that a cutoff at $M_{\rm stream}=10^{13}M_\odot$ makes only a small difference to $n(>\dot{M})$, by a factor of ~ 2 at the high-$\dot{M}$ regime corresponding to the bright SMGs. Thus, the decay of cold streams above $10^{13}M_\odot$ is not expected to alter our results in a qualitative way. On the other hand, we learn from the fact that the symbols lie far above the lower curve that the high-SFR objects at these redshifts are dominated by central galaxies in

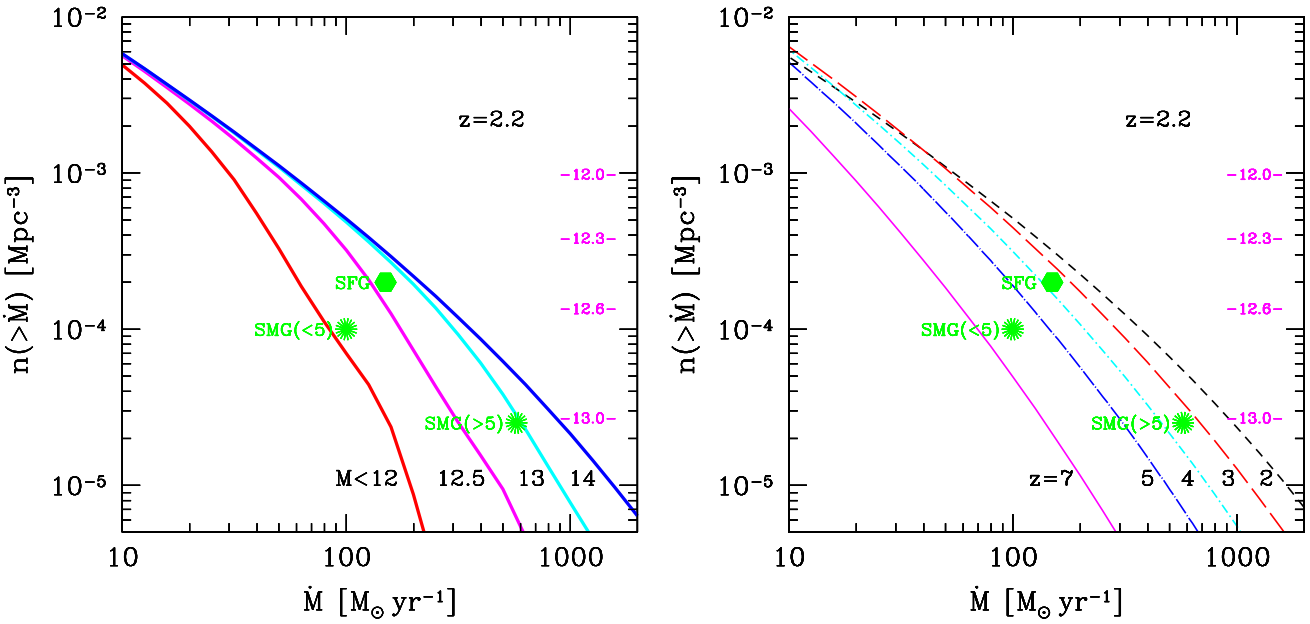

Figure 13. Comoving number density of galaxies with total gas inflow rate higher than $\dot{M}$ at $z = 2.2$, as in Fig. 4 of the main body of the Letter. The numbers on the right refer to $\log M_{\rm v}$ of haloes with the corresponding abundance. **Left:** Dependence on the maximum halo mass that contributes to cold streams, for $M_{\rm stream} = 10^{12}, 10^{12.5}, 10^{13}, 10^{14} M_\odot$. **Right:** Variation with redshift, $z = 2, 3, 4, 5, 7$.

haloes more massive than $10^{12} M_\odot$. In fact, we find that some of the SFGs and many of the bright SMGs are associated with haloes more massive than $3 \times 10^{12} M_\odot$.

Fig. 13 also shows the predicted abundance $n(>\dot{M})$ at different redshifts, now applying no finite upper mass cutoff $M_{\rm stream}$. This is justified for $z > 2$ based on our preliminary investigation of the MareNostrum galaxies at different redshifts and masses and consistent with the conjecture of DB06.[8] We see that the comoving abundance of galaxies with $\dot{M} \sim 150\, M_\odot\, {\rm yr}^{-1}$ is predicted to vary by a factor less than two between $z = 2$ and 4. By $z \sim 7$ that abundance drops by an order of magnitude. The variation with redshift is somewhat larger at the high-flux end, toward $\dot{M} \sim 10^3\, M_\odot\, {\rm yr}^{-1}$. At lower redshifts, the contribution of streams in massive haloes above $M_{\rm shock}$ is most likely overestimated by this procedure, so a similar analysis in the low-z regime should impose an upper limit at $M_{\rm stream} \simeq M_{\rm shock}$.